

Lab FAT

Machine learning

Logistic regression

Link to dataset : [<https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29>]

Anish S Ghiya

18BCB0002

Importing the dependancies

In [73]:

```
import pandas as pd
import numpy as np
import seaborn as sns # plotting
#eval metric
from sklearn.metrics import f1_score
from sklearn.metrics import roc_auc_score
from yellowbrick.classifier import ROCAUC
#model
from sklearn.linear_model import LogisticRegressionCV

#splitting train and test
from sklearn.model_selection import train_test_split
```

Understanding the dataset

(int) Age
(int) Number of sexual partners
(int) First sexual intercourse (age)
(int) Num of pregnancies
(bool) Smokes
(bool) Smokes (years)
(bool) Smokes (packs/year)
(bool) Hormonal Contraceptives
(int) Hormonal Contraceptives (years)
(bool) IUD
(int) IUD (years)
(bool) STDs
(int) STDs (number)
(bool) STDs:condylomatosis
(bool) STDs:cervical condylomatosis
(bool) STDs:vaginal condylomatosis
(bool) STDs:vulvo-perineal condylomatosis
(bool) STDs:syphilis
(bool) STDs:pelvic inflammatory disease
(bool) STDs:genital herpes
(bool) STDs:molluscum contagiosum
(bool) STDs:AIDS
(bool) STDs:HIV
(bool) STDs:Hepatitis B
(bool) STDs:HPV
(int) STDs: Number of diagnosis
(int) STDs: Time since first diagnosis
(int) STDs: Time since last diagnosis
(bool) Dx:Cancer
(bool) Dx:CIN
(bool) Dx:HPV

[illegible]

In []:

```
df.dtypes
```

Out[]:

Age	int64
Number of sexual partners	object
First sexual intercourse	object
Num of pregnancies	object
Smokes	object
Smokes (years)	object
Smokes (packs/year)	object
Hormonal Contraceptives	object
Hormonal Contraceptives (years)	object
IUD	object
IUD (years)	object
STDs	object
STDs (number)	object
STDs:condylomatosis	object
STDs:cervical condylomatosis	object
STDs:vaginal condylomatosis	object
STDs:vulvo-perineal condylomatosis	object
STDs:syphilis	object
STDs:pelvic inflammatory disease	object
STDs:genital herpes	object
STDs:molluscum contagiosum	object
STDs:AIDS	object
STDs:HIV	object
STDs:Hepatitis B	object
STDs:HPV	object
STDs: Number of diagnosis	int64
STDs: Time since first diagnosis	object
STDs: Time since last diagnosis	object
Dx:Cancer	int64
Dx:CIN	int64
Dx:HPV	int64
Dx	int64
Hinselmann	int64
Schiller	int64
Citology	int64
Biopsy	int64
dtype:	object

In []:

```
df.isnull().sum()
```

Out[]:

Age	0
Number of sexual partners	0
First sexual intercourse	0
Num of pregnancies	0
Smokes	0
Smokes (years)	0
Smokes (packs/year)	0
Hormonal Contraceptives	0
Hormonal Contraceptives (years)	0
IUD	0
IUD (years)	0
STDs	0
STDs (number)	0
STDs:condylomatosis	0
STDs:cervical condylomatosis	0
STDs:vaginal condylomatosis	0
STDs:vulvo-perineal condylomatosis	0
STDs:syphilis	0
STDs:pelvic inflammatory disease	0
STDs:genital herpes	0
STDs:molluscum contagiosum	0

```

STDs:AIDS                                0
STDs:HIV                                  0
STDs:Hepatitis B                         0
STDs:HPV                                  0
STDs: Number of diagnosis                 0
STDs: Time since first diagnosis           0
STDs: Time since last diagnosis            0
Dx:Cancer                                 0
Dx:CIN                                    0
Dx:HPV                                    0
Dx                                         0
Hinselmann                               0
Schiller                                  0
Citology                                  0
Biopsy                                    0
dtype: int64

```

In []:

```

for col in np.array(df.columns()):
    print(col)

```

```

-----
TypeError                                 Traceback (most recent call last)
<ipython-input-24-69a1d7b68ac4> in <module>()
----> 1 for col in np.array(df.columns()):
      2     print(col)

```

TypeError: 'Index' object is not callable

In []:

```

for col in np.array(df.columns):
    df.loc[df[col]=='?', col]=np.NaN

```

In []:

```

df.isnull().sum()

```

Out[]:

```

Age                                         0
Number of sexual partners                  26
First sexual intercourse                    7
Num of pregnancies                         56
Smokes                                     13
Smokes (years)                            13
Smokes (packs/year)                       13
Hormonal Contraceptives                   108
Hormonal Contraceptives (years)           108
IUD                                         117
IUD (years)                               117
STDs                                       105
STDs (number)                             105
STDs:condylomatosis                       105
STDs:cervical condylomatosis              105
STDs:vaginal condylomatosis               105
STDs:vulvo-perineal condylomatosis        105
STDs:syphilis                             105
STDs:pelvic inflammatory disease          105
STDs:genital herpes                       105
STDs:molluscum contagiosum                105
STDs:AIDS                                 105
STDs:HIV                                  105
STDs:Hepatitis B                         105
STDs:HPV                                  105
STDs: Number of diagnosis                 0
STDs: Time since first diagnosis           787
STDs: Time since last diagnosis            787
Dx:Cancer                                 0
Dx:CIN                                    0
Dx:HPV                                    0
~                                         ~

```

```

Dx                                0
Hinselmann                       0
Schiller                         0
Citology                         0
Biopsy                           0
dtype: int64

```

Fixing missing values and datatypes

In [59]:

```
df = df.fillna(0)
```

In [69]:

```

for id, dt in enumerate(np.array(df.dtypes)):
    if dt=='O':
        if np.array(df[df.columns[id]].value_counts()).shape[0] == 2:
            df[df.columns[id]] = df[df.columns[id]].astype("float")
            df[df.columns[id]] = df[df.columns[id]].astype("int")
        else :
            df[df.columns[id]] = df[df.columns[id]].astype("float")
            df[df.columns[id]] = df[df.columns[id]].astype("int")

```

In [70]:

```
df.dtypes
```

Out[70]:

```

Age                                float64
Number of sexual partners          float64
First sexual intercourse            int64
Num of pregnancies                  int64
Smokes                             int64
Smokes (years)                     int64
Smokes (packs/year)                 int64
Hormonal Contraceptives             int64
Hormonal Contraceptives (years)     int64
IUD                                 int64
IUD (years)                         int64
STDs                                int64
STDs (number)                       int64
STDs:condylomatosis                 int64
STDs:cervical condylomatosis        float64
STDs:vaginal condylomatosis         int64
STDs:vulvo-perineal condylomatosis  int64
STDs:syphilis                       int64
STDs:pelvic inflammatory disease    int64
STDs:genital herpes                 int64
STDs:molluscum contagiosum          int64
STDs:AIDS                           float64
STDs:HIV                             int64
STDs:Hepatitis B                    int64
STDs:HPV                             int64
STDs: Number of diagnosis           float64
STDs: Time since first diagnosis     int64
STDs: Time since last diagnosis      int64
Dx:Cancer                           float64
Dx:CIN                              float64
Dx:HPV                              float64
Dx                                  float64
Hinselmann                          float64
Schiller                            float64
Citology                            float64
Biopsy                              float64
dtype: object

```

In [71]:

```
df[['Hinselmann', 'Schiller', 'Citology', 'Biopsy']].values
```

Out[71]:

```
array([[0., 0., 0., 0.],
       [0., 0., 0., 0.],
       [0., 0., 0., 0.],
       ...,
       [0., 0., 1., 0.],
       [0., 0., 0., 0.],
       [0., 0., 0., 0.]])
```

Some of the rows have no values 1 like row 1 and 2 this might be an issue for softmax predictions while using MLP

In [74]:

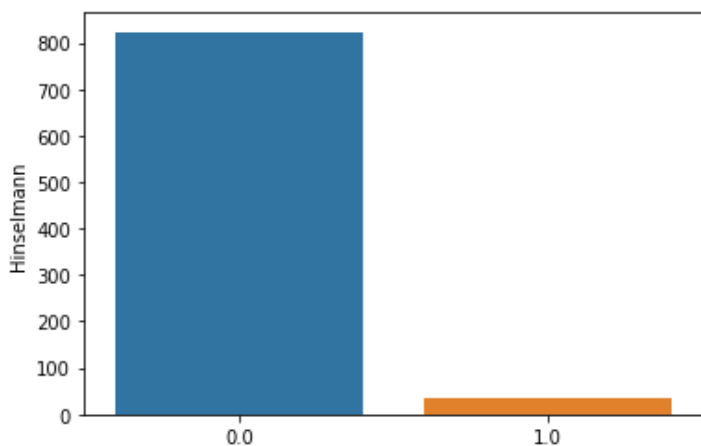
```
sns.barplot(df['Hinselmann'].value_counts().index, df['Hinselmann'].value_counts())
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

FutureWarning

Out[74]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f900b243150>



Train test split

In [79]:

```
from sklearn.model_selection import train_test_split
X = df.iloc[:, :-4]
y = df['Hinselmann']
# using stratified to ensure equal 1's in the test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42, stratify=y)
```

In [80]:

```
from sklearn.linear_model import LogisticRegressionCV
lr = LogisticRegressionCV(cv=5, random_state=0).fit(X_train, y_train)
```

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear model/ logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear model/ logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear model/ logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence

Warning: lbfgs failed to converge (status=1):

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

/usr/local/lib/python3.7/dist-packages/sklearn/linear model/ logistic.py:940: Convergence

```
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
/usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:940: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
```

In [82]:

```
lr = LogisticRegressionCV(cv=5, random_state=0, max_iter=1000).fit(X_train, y_train)
```

Evaluation metrics on test set

In [89]:

```
print("F1 Score = ", f1_score(y_test, lr.predict(X_test), average='weighted'))
print("Accuracy = ", lr.score(X_test, y_test))
```

```
F1 Score = 0.9370756915594285
Accuracy = 0.9577464788732394
```

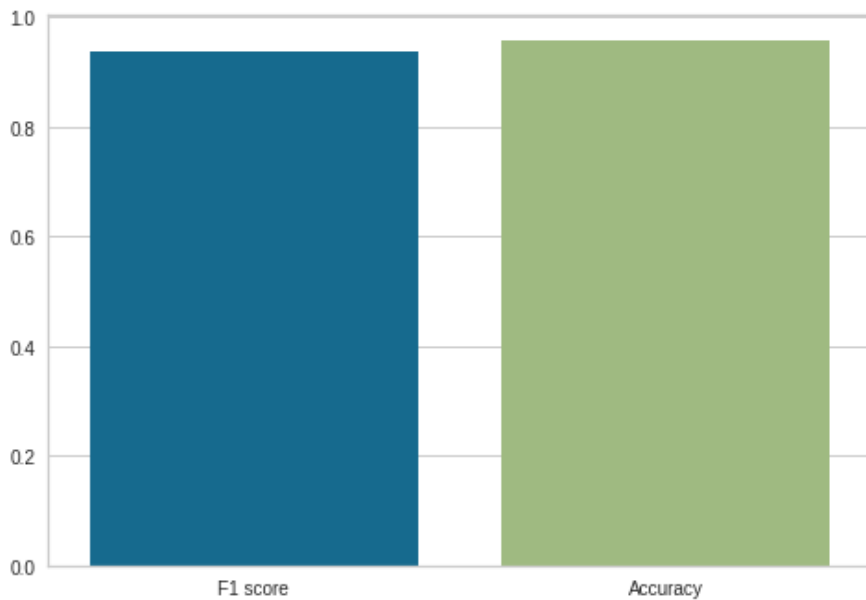
In [90]:

```
sns.barplot(['F1 score', 'Accuracy'], [0.9370756915594285, 0.9577464788732394])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the
following variables as keyword args: x, y. From version 0.12, the only valid positional a
rgument will be `data`, and passing other arguments without an explicit keyword will resu
lt in an error or misinterpretation.
FutureWarning
```

Out[90]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f8ffe249ad0>



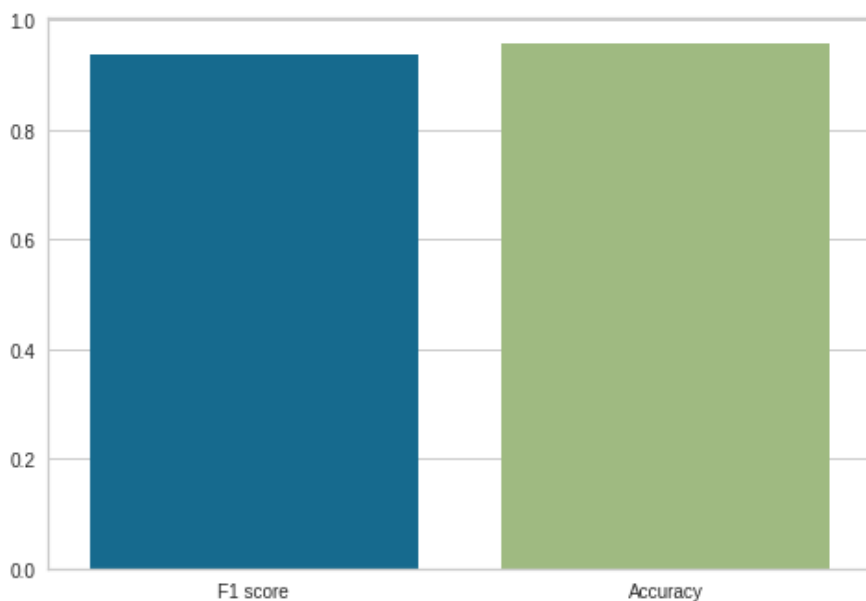
In [94]:

```
import warnings
warnings.simplefilter("ignore")
import matplotlib.pyplot as plt

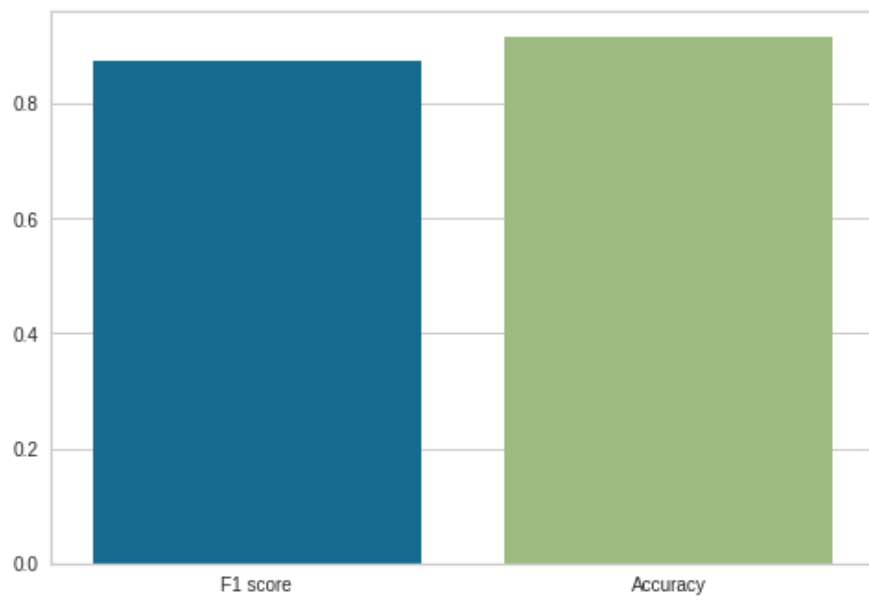
for col in ['Hinselmann', 'Schiller', 'Citology', 'Biopsy']:

    X = df.iloc[:, :-4]
    y = df[col]
    print(col)
    # using stratified to ensure equal 1's in the test set
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42, stratify=y)
    lr = LogisticRegressionCV(cv=5, random_state=0, max_iter=1000).fit(X_train, y_train)
    print("F1 Score = ", f1_score(y_test, lr.predict(X_test), average='weighted'))
    print("Accuracy = ", lr.score(X_test, y_test))
    sns.barplot(['F1 score', 'Accuracy'], [f1_score(y_test, lr.predict(X_test), average='weighted'), lr.score(X_test, y_test)])
    plt.show()
```

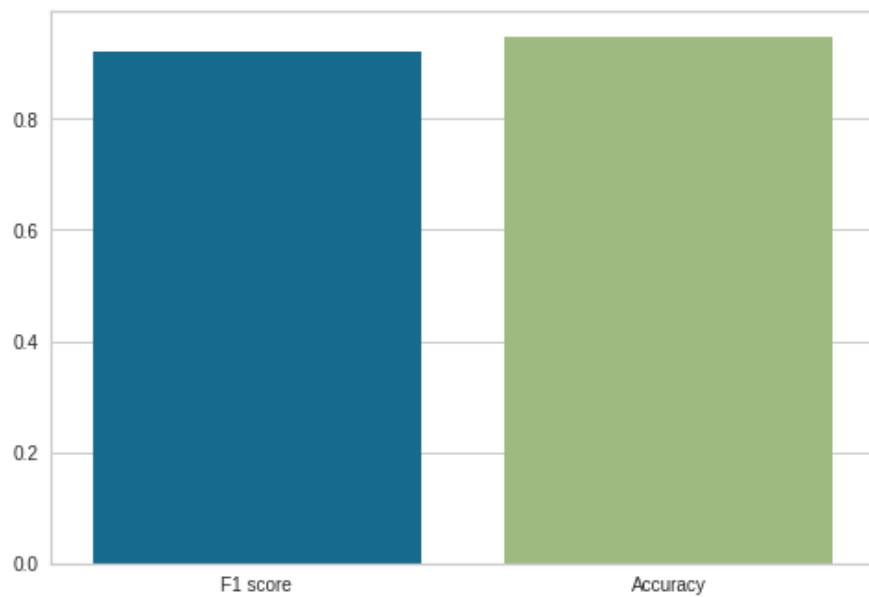
Hinselmann
F1 Score = 0.9370756915594285
Accuracy = 0.9577464788732394



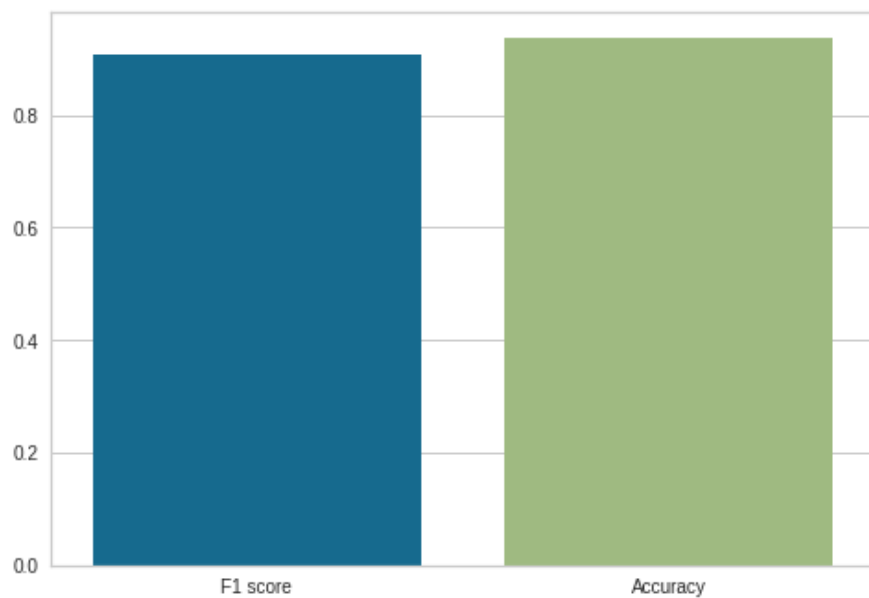
Schiller
F1 Score = 0.8751035625517812
Accuracy = 0.9154929577464789



Citology
F1 Score = 0.9214909711433155
Accuracy = 0.9471830985915493



Biopsy
F1 Score = 0.9059667093469911
Accuracy = 0.9366197183098591



In [98]:

from sklearn.model_selection import GridSearchCV

```
from sklearn.model_selection import GridSearchCV
log_params = {'penalty':['l1', 'l2', 'elasticnet'],
              # 'C': [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 100],
              # 'solver':['liblinear', 'saga']
              'solver': ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'],
              # 'multi_class':['auto', 'ovr', 'multinomial']
              }
# log_model = GridSearchCV(LogisticRegression(), log_params, cv=5)
log_model = GridSearchCV(LogisticRegressionCV(cv=5, random_state=0, max_iter=1000).fit(X
_train, y_train), param_grid = log_params, scoring = 'accuracy')

log_model.fit(X, y)
print(log_model.best_params_)
print(log_model.best_score_)
```

```
{'penalty': 'l1', 'solver': 'liblinear'}
0.9358969128246974
```

In []: